

Special Section

# Beyond the Gallery Walls: Tools and Methods for Leading End-Users to Collections Information

by Erin Coburn and Murtha Baca

Erin Coburn, data standards administrator at the J. Paul Getty Museum, can be reached at [ecoburn@getty.edu](mailto:ecoburn@getty.edu).

Murtha Baca, head, Getty Vocabulary Program and Digital Resource Management, Getty Research Institute, can be reached at [mbaca@getty.edu](mailto:mbaca@getty.edu).

The library and archive communities have a long-standing history of organizing and managing their information in a way that facilitates access to their holdings, both within their home institutions and via union catalogs and consortial bibliographic utilities such as the Research Libraries Information Network (RLIN) and OCLC's *WorldCat*. Metadata element sets and information protocols, such as MARC and the Encoded Archival Description (EAD), and controlled vocabularies and authority files, such as the Library of Congress name and subject authorities, the *Thesaurus for Graphic Materials* and the *Art & Architecture Thesaurus* (AAT), are the data standards and structures that have become synonymous with the way that the library and archival communities provide access to information about their holdings. In the traditional library world, for instance, the AAT is typically used for data values in the form/genre field in a MARC record when cataloging rare books and special collections materials.

### Museum Collections: Making Information Available

Museums have only relatively recently come to an awareness that the approach that libraries and archives have taken for decades is also essential for making information on their own collections available. Museum information has a history of being hoarded if not outright hidden in curatorial files. The emerging use of computerized collections management systems in museums caused many curatorial departments to relinquish control over information about their collections, but in many cases the information was limited to whatever was deemed necessary for proper maintenance and inventory. With the rise of the World Wide Web, and the demands of users (and museum trustees)

who expect to have collections information one click away, museums have begun to take seriously the audience of users who may want access to their collections even if those users will never physically visit their museums. Meeting this demand is not as simple as just digitizing a collection and creating a link to the image repository from an institution's home page. It is no simple task to ensure that users whose profiles, needs and information-seeking behaviors are many and varied will be successful in finding what they are looking for.

A good starting point for museums grappling with how to enable diverse audiences to effectively access information about their collections is to look within the walls of the museum itself. Collections management systems are growing in their capacity for recording and documenting collections. Cataloging is no longer limited to a registrar entering core information about a work of art upon acquisition. Other key information stakeholders are now using collections management systems to record information on conservation treatments, exhibitions, provenance, published sources, rights and so forth. As a result, the range of users within a museum who need access to – and who help to create – collections information is growing to include curators, conservators, educators and even docents.

This expanded role and importance of collections management systems as the source for gathering a wide range of diverse information about the objects in a collection is resulting in a new way of thinking for museums. Collections management systems are becoming collections *information* systems, and documentation is going beyond just registration; it is serving the larger purpose of aggregating all relevant information about the works in a collection and preparing that information for



delivery or publication in a variety of environments and to a variety of users, both internal and external.

In addition to the Web, another area that museums are exploring as a means to providing greater access to their collections is with public access or kiosk systems. Until a decade or so ago, many museums and cultural heritage institutions packaged selective information about their collections on a CD-ROM, which was then made accessible from a kiosk station within the museum itself. This kind of closed, hard-coded system also gave museums a product to sell; but updating the CD-ROMs to account for changes in attribution or to include new acquisitions was so costly as to be infeasible.

Museums are now creating models and processes in which a collections information system serves as the starting point for publishing selective information to a content management system, or a set of content management and manipulation tools that make it possible for data to be edited and enhanced to better fit the needs of the audience(s) using the public access system or the institution's Web pages. The benefit of this model is that it creates the opportunity to change, update and add additional information about the collection in a dynamic way, rather than producing CD-ROMs, which, for many if not most institutions, become out of date the moment they are produced.

### Creating Access: Two Misconceptions

Creating access to a museum's collection on the Web opens up a whole new level of complexity that requires careful thought, for users can no longer be simply identified as museum staff or museum visitors, nor can their needs be neatly categorized. Making collection information available on the Web means that everyone with access to a computer that is connected to the Internet is a potential user, regardless of age, educational and cultural background or native language.

**The "project" approach.** There are two common misconceptions that many museum professionals have about making their information available to larger audiences and both have the potential to create significant setbacks for advancing the ability of museums to better share and disseminate information on their collections. The first is that making collections available on line or in a public access system is a discrete, finite project that will sooner or later come to an end. In most cases, this approach involves seeking grant money or other sources of financial support in order to fund the project. In recent years, many museums have made their collections, or at least portions of them, available on line. While grants have played a large part in making this possible and this is a pragmatic approach given the limited financial resources available to most cultural heritage institutions, one negative aspect of this approach is that many museums tend to see the process of making their collections accessible on the Web or in kiosk systems as "one-off" projects, rather than as an ongoing part of their core mission and activities.

When museums take this kind of approach, the lion's share of the resources tend to go to the activities strictly related to creating digital images of collection objects, while activities such as creating additional content and contextual tools, building access points based on data standards and controlled vocabularies, and identifying audiences and their needs and behaviors are all relegated to "phase two" or "we'll deal with that later." The authors of this article can't count the number of times that, after trying to convince colleagues from other institutions that they need to invest time and labor in building good data and implementing standards and controlled vocabularies on the back end, they have received the response, "We don't have time to deal with that now." A museum might be able to boast that it has its entire collection of art works available on the Web and to say to the philanthropic agency that funded the project that they have fulfilled their grant by making their collection accessible on line, but what exactly is meant by *accessible*? Can a teacher preparing a class on women artists find all the works in the collection by female artists? Can a scholar or researcher find the history of the ownership of a particular painting in the collection? Can a curator from another institution identify the 17th-century drawings by French artists in the collection without calling the museum's curator of drawings? Can a student find the ritual vessels depicting warriors in the collection? For the purposes of this paper, we are defining access as "the ability of a user to find and use information in electronic form." (See also Joan M. Reitz's excellent *Online Dictionary of Library and Information Science* for definitions of *accessibility*, *authorities*, etc.)

**Staffing requirements.** The second misconception about the work involved in making collections information available to larger audiences is the notion that this can be an added responsibility to the registrar's position or to some other position within the institution. While smaller institutions may have no other choice than to give this extra set of tasks to the registrar, institutions need to start facing the fact that the cluster of activities related to creating, managing and publishing collections information will require new people, new skill sets and, yes, new positions. Increasingly, as Ken Hamma has also noted in his article in this issue, museum positions include job titles such as *Data Standards Administrator* (J. Paul Getty Museum) and *Manager for Information Standards* (Metropolitan Museum of Art) and departments such as *Collections Information and Access* (San Francisco Museum of Modern Art). To better understand the scope of the responsibilities associated with these positions and how they differ from the role of a registrar, we need look no further than the job descriptions.

The National Academy of Design in New York has a registrar position posted on *Aviso* (the monthly newsletter of the American Association of Museums) seeking

an experienced, full-time registrar responsible for all aspects of registration, including accessioning, incoming and outgoing loans, traveling exhibitions, docu-

The goal that underlies these new kinds of positions is to make collections information more readily accessible for a variety of audiences and uses.

ment management, and inventory maintenance. The registrar is responsible for maintaining the inventory control for all objects in the museum's collection and an accurate record of the location of each object in the museum; arranging the packing, transportation, and insuring of all objects passing into and out of the museum; and coordinating the de-installation of all objects for exhibition.

The new jobs that deal with the larger scope of museum information have resulted in positions such as *Director of Metadata and Cataloging* at ARTstor (an initiative to make cultural heritage digital images available for educational purposes), who is tasked to

develop metadata standards for a wide range of collections; develop guidelines and document best practices for the development of metadata; analyze and evaluate metadata in collections; oversee and coordinate the work of catalogers creating, revising or enhancing metadata and create a quality control process for such work; collaborate with other staff in the development of the database architecture and features of the software environment; and serve as the principal liaison to metadata standard developers and cataloging authorities in the educational, cultural and scholarly communities.

While ARTstor is not a museum (but does – or should – depend on museums for a good portion of its virtual collections), the scope of this position could easily be applied to museums with diverse collections or to institutions managing art collections as well as collections that fall under archives and library holdings.

## The Role of Standards

The goal that underlies these new kinds of positions is to make collections information more readily accessible for a variety of audiences and uses. Data and metadata standards and controlled vocabularies are crucial in making this possible. There are three types of standards pertinent to the proper management of museum information: data structure standards, data value standards and data content standards or guidelines for how to format the data values that are used to populate the data structures.

**Data structure standards.** Data structure standards often take the form of metadata element sets. *Categories for the Description of Works of Art* developed by the Art Information

Task Force under Getty sponsorship and the Visual Resources Association's *VRA Core Categories* are two widely used metadata element sets for describing cultural objects and their images. Fortunately for museums, the use of collections management software typically provides a structure for managing and documenting objects in an organized and meaningful way. But simply having a collections management system will not guarantee easy access to the works in a museum or other cultural heritage repository.

**Data value and data content standards.** The words and terms (known as data values in the information science realm) that a museum uses to document works of art are of great importance. Once again, the cultural heritage community is fortunate to have a variety of tools, in the form of controlled vocabularies, classification systems and thesauri, to help with the proper selection of data values for cataloging. These include, but are certainly not limited to, the *Art & Architecture Thesaurus*, *Union List of Artist Names* (ULAN), *Getty Thesaurus of Geographic Names* (TGN) and of course the *Library of Congress Authorities*.

There are multiple uses of controlled vocabularies and thesauri, and their importance in creating consistent documentation and facilitating end-user access to museum information cannot be overestimated. Vocabularies and classification systems can serve as research tools to help clarify the proper meaning or use of a term. For example, a cataloger providing subject access to a work of art can refer to the ICONCLASS system (a subject classification system for iconographic research and the documentation of images) to clarify that a wheel is the identifying attribute for Saint Catherine of Alexandria, not Saint Catherine of Siena.

Vocabularies can also serve as authorities to which collections management systems can link in order to help properly identify the scope and meaning of terms applied to a work of art and to establish preferred terms or headings. For example, a Flemish manuscript from Bruges can be cataloged with a link to the TGN as a geographical place authority, which would preserve the meaning of Bruges as being part of the historical region of Flanders in present-day Belgium. These two uses of vocabularies and thesauri (as authorities and as lookup tools) have been widespread for decades in the library and archive communities, and in more recent years, they have been implemented increasingly in the museum community as well. A third way to use vocabularies and thesauri, and one that is of great interest and potential benefit to any institution seeking to enhance end-user access to its collections on line, is as "searching assistants."

**Best practices.** Finally, good descriptive cataloging requires knowledge of best practices for museum documentation and the ability to implement those practices within an institution; this is where the data content standards or guidelines come in. As mentioned above, it is now common to find a variety of staff involved in entering descriptive data about a collection in a museum's database. Ensuring that this level of cataloging across departments and by diverse staff adheres to a set of guidelines is crucial for creating consistent, accurate access to collections. In the museum world, standards for the selection, organization and formatting of content have only relatively recently begun to get the attention they so desperately need.

A new initiative to help the museum community address these issues is *Cataloguing Cultural Objects* (CCO), a Visual Resources Association project that seeks to provide standards for data content in cataloging cultural works and their images. But we must repeat that the key element with all standards and guidelines is the ability for museums to have the proper people in place to implement and support them in ways that are part of the normal workflow of the institution.

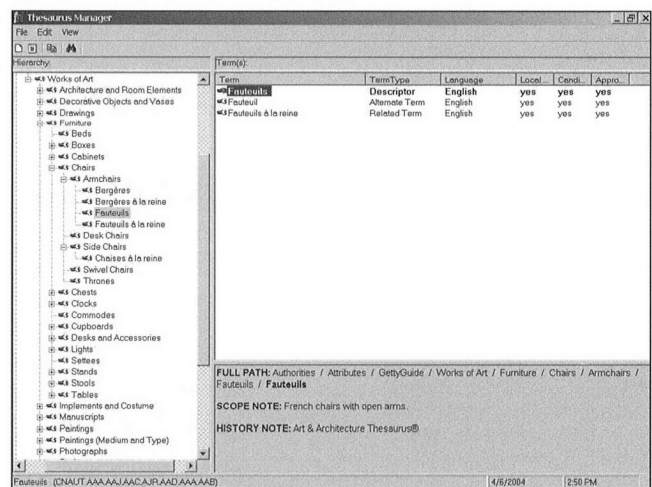
## Re-Purposing Data from Collection Information Systems: Projects at the Getty

Adherence to metadata standards, using controlled vocabularies and following best practices for descriptive cataloging open up numerous possibilities for re-purposing data and increasing access to collections for diverse audiences with varying needs. Access to collections should be data-driven – ideally coming from a central repository, such as a collections management system, that is maintained by key information stakeholders in accordance with a set of accepted rules and practices. This model enables the institution to efficiently and effectively manage, maintain and preserve data over time.

The Getty Museum has made it a policy to re-purpose data from its collections information system to enhance access to its collection for a variety of users, internal and external; as much as possible, the content we deliver to our users on the Web, in our kiosk system and in our hand-held system is “fed” by data from our collections information system.

The Getty Museum builds collection-specific thesauri incorporating terminology from existing vocabularies such as the AAT, in addition to including “local” headings that help to categorize the objects in its collection in meaningful ways for non-expert users. One of the museum's local thesauri is a vocabulary of object types represented in the museum's collection. A section, or facet, of this thesaurus deals with furniture in the collection, which is then broken down into categories such as *chairs*, and further by types of chairs such as *fauteuils* (French chairs with open arms) (Figure 1). In the museum's database, objects are linked to their appropriate object name in the thesaurus. This data is then re-purposed on the Web in two different ways.

First, the structure of the thesaurus itself assists users in

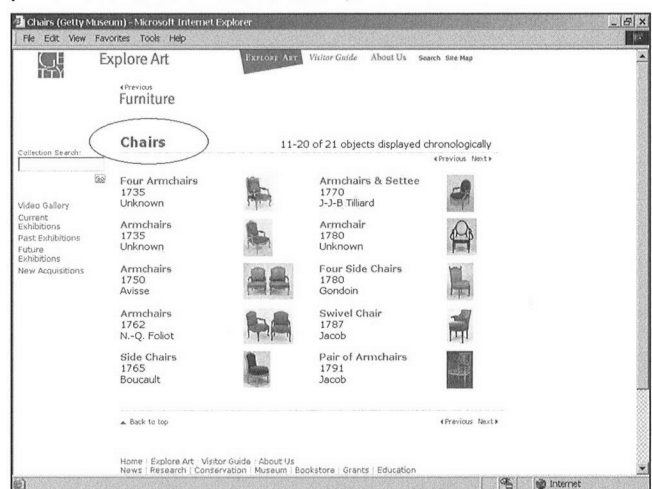


**FIGURE 1.** View of the Getty Museum's collection-specific thesaurus in its thesaurus manager application, a part of the museum's collections information system. © J. Paul Getty Trust.

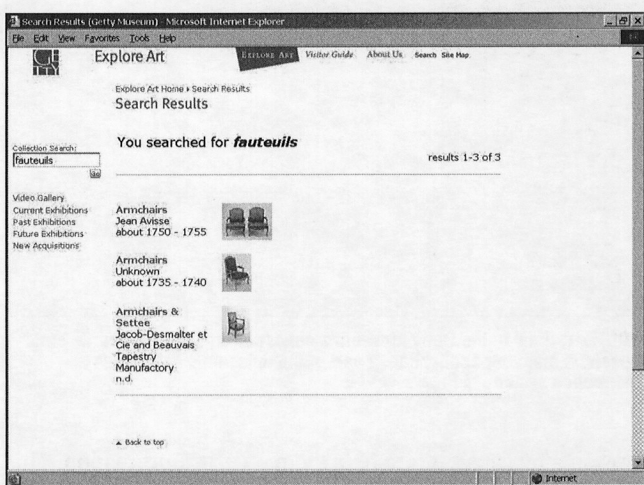
browsing by categories to help them retrieve information. This is particularly helpful for users who are unclear about what exactly they are searching for or for users who may not be familiar with art terminology. The Getty Museum exposes the broader, more generic categories of the thesaurus to facilitate retrieval for these types of users. For example, someone interested in chairs could select this heading as a category and retrieve results immediately without having to be confronted with a list of specific types of chairs, such as *fauteuils* (Figure 2).

Secondly, the Getty Museum utilizes the specific terms from the object type thesaurus that are linked to the objects in the database to assist users who know precisely what they are interested in finding. These data values or names are

**FIGURE 2.** Screen capture of the Getty Museum's category for *Chairs* as part of its collection on line. © J. Paul Getty Trust.



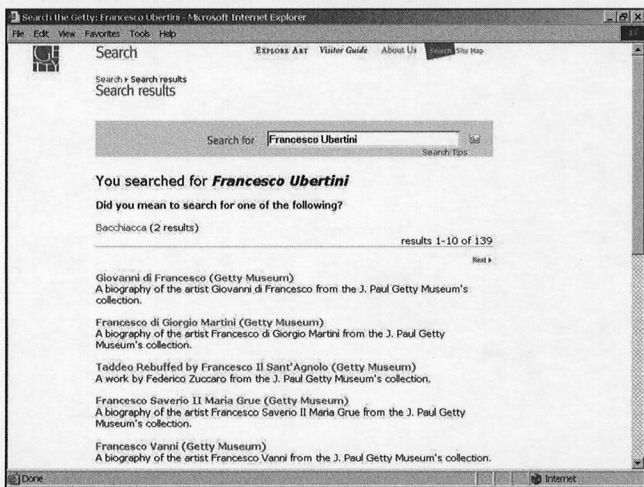
One of the benefits of vocabulary-assisted searching is that it frees catalogers and other data creators from having to enter the many variants or alternate names commonly attributed to objects and artists...



**FIGURE 3. Results of a search on *fauteuils* performed on the Getty Museum's collection on line. © J. Paul Getty Trust.**

brought over to the META keyword tag for the individual object records available in the Getty Museum's collection on the Web. The META keyword tag is for select words and phrases that properly describe the Web resource and are used by a variety of search engines, including the Getty's own site-wide search engine, to facilitate retrieval. For example, if a curator from another museum were interested to know if the Getty had any *fauteuils* in its collection, he or she could perform a

**FIGURE 4. Screen capture of a search for *Francesco Ubertini* on the Getty's site-wide search engine. © J. Paul Getty Trust.**



keyword search on *fauteuils* and would retrieve three results (Figure 3). The first two results are retrieved not because the term *fauteuils* appears explicitly on the text of the Web pages for these objects, but because the term *fauteuils* from the local thesaurus has been embedded in the META keyword tags for these Web pages and is indexed by the site-wide search engine.

The Getty has also begun to integrate controlled vocabularies with its internal search engine to test the vast possibilities of increasing access to collections with vocabulary-assisted searching. As of this writing (April 2004), the Getty has implemented the ULAN as a searching assistant on its site-wide search engine. One of the benefits of vocabulary-assisted searching is that it frees catalogers and other data creators from having to enter the many variants or alternate names commonly attributed to objects and artists; as long as one of

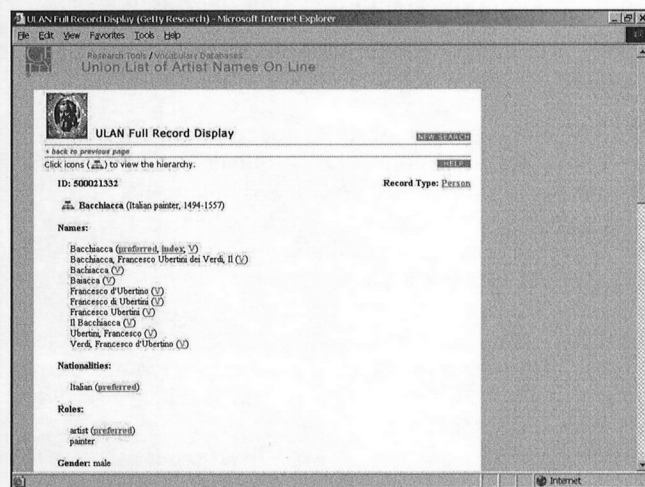


FIGURE 5. ULAN record for the artist *Bacchiacca*. © J. Paul Getty Trust.

the values or words used to describe a particular work or creator is represented in the vocabulary to which the system is linked, the thesaurus will take care of proper retrieval.

For example, a user performing a search on the Getty's website for works or information about the artist *Francesco Ubertini* will obtain results with a prompt asking, "Did you mean to search for one of the following? *Bacchiacca* (2 results)" (Figure 4). What happens behind the scenes is that the search engine first looks for an exact match on the Getty's Web pages and finds none. It then runs the name *Francesco Ubertini* against the ULAN, where it finds a match in a variant name on one of the clusters of names that form the ULAN records (Figure 5). All of the names for this artist, preferred

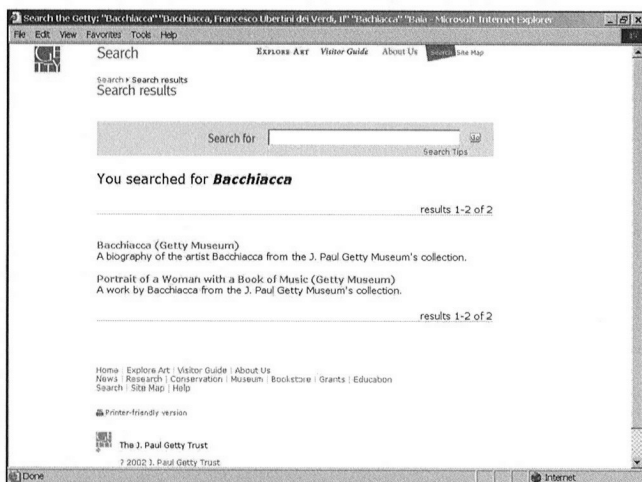


FIGURE 6. Screen capture of a ULAN vocabulary-assisted search result on the Getty's site-wide search engine. © J. Paul Getty Trust.

and variant, are then submitted to the search engine which results in a match on *Bacchiacca* in the Web pages for the museum's collection. This whole process is hidden from the user. The user's verbatim search results in Getty Web pages that contained either the words *Francesco* or *Ubertini*.

However, by giving the user the opportunity to select the ULAN record for *Bacchiacca*, which is the name for Francesco Ubertini that appears in most of the scholarly literature, the user is presented with two *exact* matches from the museum's collection (Figure 6); one a link to the biography of the artist and the other a link to a painting attributed to him.

The Getty Museum is just one example of a museum that is reaping the benefits of managing the data about its collection in a way that is standards-compliant and utilizes controlled vocabularies. The practice of instituting and implementing a model for managing data has led to the ability of museums to more easily migrate their data to new systems, to publish their data in a dynamic fashion to websites or public access systems and to be able to contribute their collections information to consortial initiatives such as Museums and the Online Archive of California (MOAC) and the Cultural Materials project of the RLG (Research Libraries Group). Museums still have a long way to go before they achieve what libraries and archives have been doing for decades to facilitate access to their holdings. But if we continue to implement standards and best practices for data-driven publishing of our collections, it won't be long before the library, archive and museum communities can create successful models of interoperability and integrated access.

#### Discussion of metadata schemas, controlled vocabularies and collection-specific thesauri for art and material culture:

Baca, M. (2003). Practical issues in applying metadata schemas and controlled vocabularies to cultural heritage information. *Cataloging & Classification Quarterly* 36(3-4), 47-55.

Baca, M. (In press). Fear of authority? Authority control and thesaurus building for art and material culture information. *Cataloging & Classification Quarterly* 38(3-4).

#### Resources mentioned in this article:

**AAT** (Art & Architecture Thesaurus) [www.getty.edu/research/conducting\\_research/vocabularies/aat/](http://www.getty.edu/research/conducting_research/vocabularies/aat/)

**ARTstor** [www.artstor.org/](http://www.artstor.org/)

**Aviso** [www.aam-us.org/aviso/index.cfm](http://www.aam-us.org/aviso/index.cfm)

**CCO** (Cataloguing Cultural Objects) [www.vraweb.org/CCOweb/index.html](http://www.vraweb.org/CCOweb/index.html)

**CDWA** (Categories for the Description of Works of Art) [www.getty.edu/research/conducting\\_research/standards/cdwa/](http://www.getty.edu/research/conducting_research/standards/cdwa/)

**EAD** (Encoded Archival Description) [www.loc.gov/ead/](http://www.loc.gov/ead/)

**ICONCLASS** <http://iconclass.nl/index.html>

**Library of Congress Authorities** <http://authorities.loc.gov/>

**MARC** (MAchine-Readable Cataloging) [www.loc.gov/marc/](http://www.loc.gov/marc/)

**MOAC** (Museums and the Online Archive of California) [www.bampfa.berkeley.edu/moac/](http://www.bampfa.berkeley.edu/moac/)

**Online Dictionary of Library and Information Science** (Joan Reitz) [www.wcsu.edu/library/odlis](http://www.wcsu.edu/library/odlis)

**RLG Cultural Materials** (Research Libraries Group) <http://culturalmaterials.rlg.org/>

**RLIN** (Research Libraries Information Network) [www.rlg.org/rlin.html](http://www.rlg.org/rlin.html)

**TGM** (Thesaurus for Graphic Materials) [www.loc.gov/rr/print/tgm1/](http://www.loc.gov/rr/print/tgm1/)

**TGN** (Getty Thesaurus of Geographic Names) [www.getty.edu/research/conducting\\_research/vocabularies/tgn/](http://www.getty.edu/research/conducting_research/vocabularies/tgn/)

**ULAN** (Union List of Artist Names) [www.getty.edu/research/conducting\\_research/vocabularies/ulan/](http://www.getty.edu/research/conducting_research/vocabularies/ulan/)

**VRA Core Categories** (Visual Resources Association) [www.vraweb.org/vracore3.htm](http://www.vraweb.org/vracore3.htm)

**WorldCat** (OCLC) [www.oclc.org/worldcat/default.htm](http://www.oclc.org/worldcat/default.htm)